

Visualising energy data

Diane Donovan*

Trevor Pickett*

Richard Wilson*

(Received 12 November 2004, revised 7 July 2005)

Abstract

The Australian energy market is in the final stages of deregulation. These changes have created a dynamic environment which is highly volatile and competitive with respect to both demand and price. Our current research seeks to visualise aspects of the National Energy Market with a view to developing techniques which may be useful in identifying significant characteristics and/or drivers of these characteristics. In order to capture the complexity of the problem we explore a suite of different visualisation techniques, which, when combined into a unified package, highlight aspects of the problem. The particular problem visualised here is “Does the data exhibit characteristics which suggest that the time of day, day of the week, or the season, affect the variation in demand and/or price?”

*University of Queensland, Brisbane, AUSTRALIA. <mailto:dmd@maths.uq.edu.au>

See <http://anziamj.austms.org.au/V46/CTAC2004/Don2> for this article, © Austral. Mathematical Soc. 2005. Published September 22, 2005. ISSN 1446-8735

Contents

1	Introduction	C957
2	Relevant background	C958
3	The data	C959
4	Visualisation	C962
5	A final visualisation	C966
6	Conclusion	C968
	References	C969

1 Introduction

The main purpose of this study is to investigate the predictive powers of various visualisation techniques in identifying significant characteristics within the given data set and/or drivers of these characteristics. The physical setting for this study is the National Electricity Market, and we investigate the conformance of participants during varying time frames.

The energy industry is of interest as there is a dynamic interplay between production and use, transfer and storage, and purchase and sale, and when these factors are coupled with issues of weather and technological advances they play a major role in driving the market. The complexity of these underlying forces together with vast amounts of data make it difficult to establish long term cycles within the market and thus the provision of useful visualisation requires data simplification and careful variable selection.

Much of the existing analysis of the data is conducted on half hourly

intervals over an entire day, and across all 365 days of the year. By utilizing various visualisation techniques we show that the data exhibits characteristics which are relatively constant for certain time frames in any given day and that there appears to exist characteristics which change over a twelve month period. However, the analysis presented here does not reveal characteristics which occur over a seven day cycle and we anticipate that further refinement of the data will be required (such analysis is part of the ongoing work). An important aspect here is the use of different visualisation techniques to highlight varying characteristics within the data. For additional analysis of the use of visual methods for the identification of characteristics in data sets see [6, 3] and for an alternative approach to the visualisation of the spot price for the Australian Energy Market see [1].

To ensure subsequent discussions are meaningful Section 2 describes relevant aspects of the Australian energy market. Section 3 then discusses details of the relevant data sets. A suite of visualisation techniques have been utilised at various times and are discussed in Sections 4 and 5.

2 Relevant background

The National Electricity Market (NEM) was established in December 1998 as part of a process to deregulate Australian energy markets. The aim was to provide for the wholesale supply and purchase of electricity in five Australian states and territories; New South Wales, Victoria, Queensland, South Australia and the Australian Capital Territory. The NEM supplies energy to 7.7 million Australian customers and trades approximately 8 billion dollars of energy per year. The market is based on an open access grid consisting of transmission and distribution networks across all states and territories. It seeks to promote a competitive environment at each stage in the electricity production and supply chain.

Currently the NEM is divided into five interconnected regions, which are

the above five states, with the ACT incorporated into NSW and the Snowy designated as a fifth region. Each region has a regional reference node which is usually a major city or generating centre. Generally all electricity in the NEM is traded through a spot market managed by the National Electricity Market Management Company Limited (NEMMCO). This spot market provides the interface between generating operators and electricity retailers. Generating operators provide bids, comprising of price and associated quartiles, and offer to supply NEMMCO differing amounts of energy at given prices. In general, generators sell electricity into the pool, and retail and wholesale users pay for the electricity they draw from the pool. NEMMCO ensures that there is enough electricity to meet demand. For each reference node, in each of the five regions, NEMMCO uses the bids and constraints, such as interconnecting capacities and energy transmission and distribution losses, to set a Regional Reference Price (RRP). Based on the combined information NEMMCO centrally co-ordinates the dispatch process, and schedules generators by sending generating instructions every five minutes. Decisions are made on the principle of determining the most efficient way to meet demand.

3 The data

NEMMCO's website [4] provides the general public with access to NEM data. For each region we have accessed data on Total Demand, TD, in megawatts, and the Region Reference Price, RRP, in dollars per megawatt hour. NEMMCO provides an online repository of this data, available in 30 minute intervals from October 13, 1998. We have sought to compare the change in relative demand and price for Queensland against the change for a combination of other NEM regions.

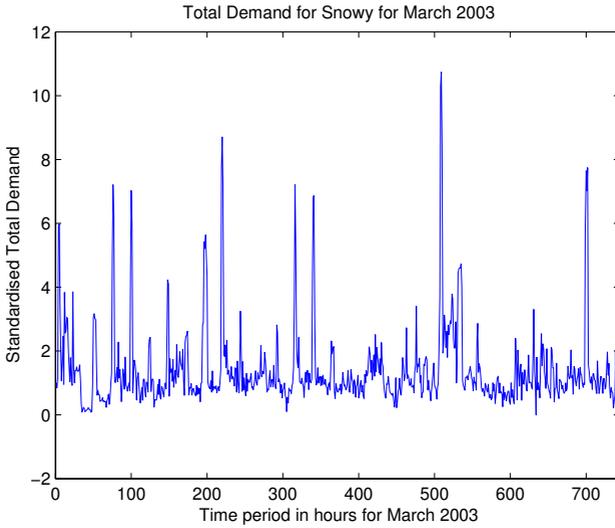
Before visualisation could be possible the data must be massaged into a workable format. The 48 half hourly time steps are judged to be too

high resolution for our requirements, so the data is averaged over two half hourly time steps, before and on each hour of the day; that is, we average $TD_{i-0.5}$ and TD_i ($RRP_{i-0.5}$ and RRP_i), where TD_i (RRP_i) represents the TD (RRP) on the hour i , for $1 \leq i \leq 24$.

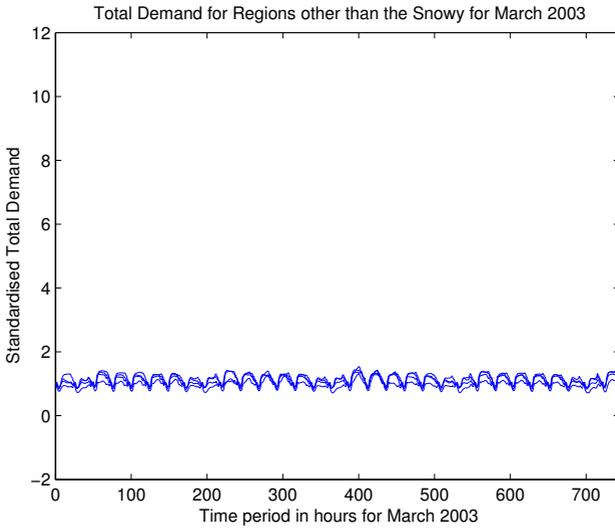
Next, as recommended by Dywer [2], the TD is standardized to ensure that meaningful comparisons are conducted across regions. So, for each region the TD at 1 a.m. on each day is a reference value scaled to 1. Each subsequent time step for that day was calculated as a ratio of the reference value. Such a modification is justified since we wish to visualise the change from one time step to the next across regions, not the difference in raw TD data. Note that we may have chosen any given time step as the reference value, provided we choose the same time step for all regions. This standardization process was not applied to the RRP data, as we wish to consider the difference in price per megawatt hour.

The visualisation displayed in Figure 1 provides a snapshot of the data for TD for the month of March. It indicates that TD for Queensland, NSW, Victoria and South Australia have a high degree of similarity, while the Snowy region's data does not. This relationship is typical for all months of 2003 and typical for the RRP data. In addition, an analysis of the log of the data entries once again shows there is a high degree of variation between the Snowy and the other regions. Therefore, in order to obtain an accurate estimate of characteristics in Queensland relative to other regions, we exclude the Snowy data.

To investigate the behaviour of a single region versus the other three regions (in this case Queensland versus NSW, Victoria and South Australia) the average of the TD (RRP) for the three regions NSW, Victoria and South Australia is taken for each time step, then subtract this value from the corresponding Queensland value. This provided us with an estimate of how Queensland differed from the average of the selected regions. The TD and RRP data is stratified into nine bands as set out in Table 1. The bands were selected to distribute the data into discrete (and visually distinct) classes.



(a)



(b)

FIGURE 1: A comparison of the TD data: (a) Snowy region; (b) other regions.

TABLE 1: Band table TD and RRP data.

Total demand data x_1	RRP data x_2	Band
$x_1 > 0.2$	$x_2 > 100$	4
$0.2 \geq x_1 > 0.15$	$100 \geq x_2 > 4$	3
$0.15 \geq x_1 > 0.1$	$4 \geq x_2 > 2$	2
$0.1 \geq x_1 > 0.05$	$2 \geq x_2 > 1$	1
$-0.05 \leq x_1 \leq 0.05$	$-1 \leq x_2 \leq 1$	0
$-0.1 \leq x_1 < -0.05$	$-2 \leq x_2 < -1$	-1
$-0.15 \leq x_1 < -0.1$	$-4 \leq x_2 < -2$	-2
$-0.2 \leq x_1 < -0.15$	$-100 \leq x_2 < -4$	-3
$x_1 < -0.2$	$x_2 < -100$	-4

The upper and lower classes partition off the very high and very low values in the data, and the intermediate classes are intended to provide a reasonable distribution of the remaining data. However, note there is much more variation in the RRP values, and so checks were carried out to ensure the banded data reflected the raw data.

The data was stored in either twelve ($s_i \times 24$) arrays m_i for $i = 1, \dots, 12$, and $s_i \in \{28, 29, 30, 31\}$, corresponding to each month, or in a 365×24 array containing the data for an entire year (366×24 in a leap year).

4 Visualisation

The visualisation process began with the construction of a 3-D surface plot for TD and RRP for each of the twelve arrays m_i . The difference between Queensland's TD (RRP) and the average of the three selected regions was sorted into nine bands from -4 to 4 and was mapped against the vertical axis. Thus the entry z in cell (x, y) represented the value in cell (x, y) of the

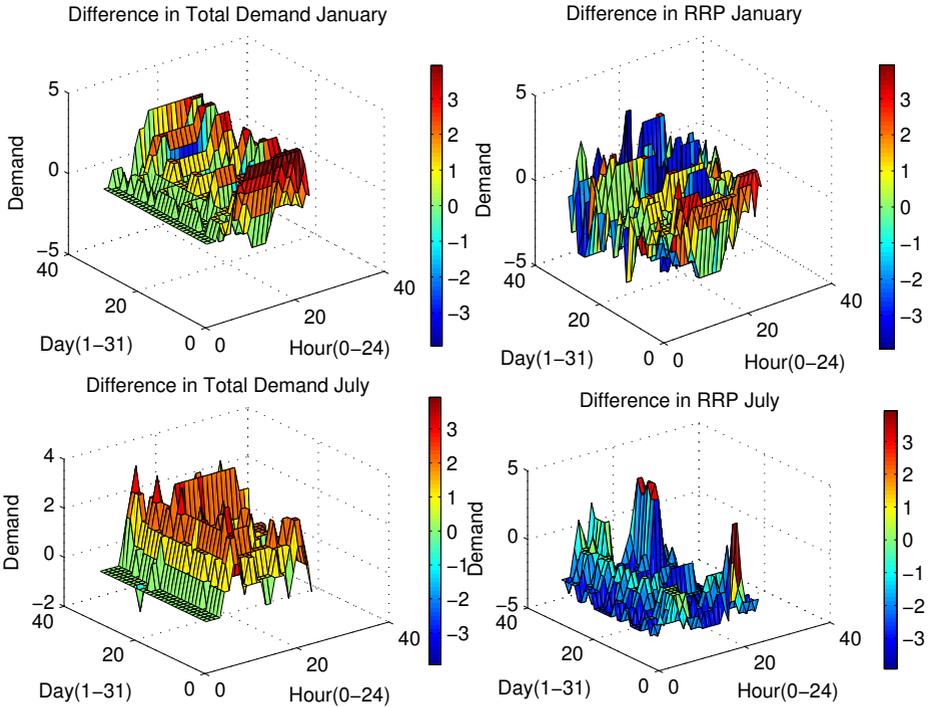


FIGURE 2: The difference in TD and RRP for Queensland against the other 3 regions for January and July 2003.

matrix m_i . The visualisation displayed in Figure 2 gives the corresponding surface plots for January and July, 2003, see also [5].

The TD plots tend to indicate the presence of two separate characteristics. First, see that for all days y there is a strong tendency for the value in cells (x, y) , for $x = 0, \dots, 5$, to be within the first band. Thus for the first five hours of each day Queensland's TD is close to the average TD for the three selected regions. However, this characteristic is not as prominent in the RRP data. Secondly, for both sets of data there appears to be troughs in the surface plot which may indicate days when Queensland's TD and RRP is uni-

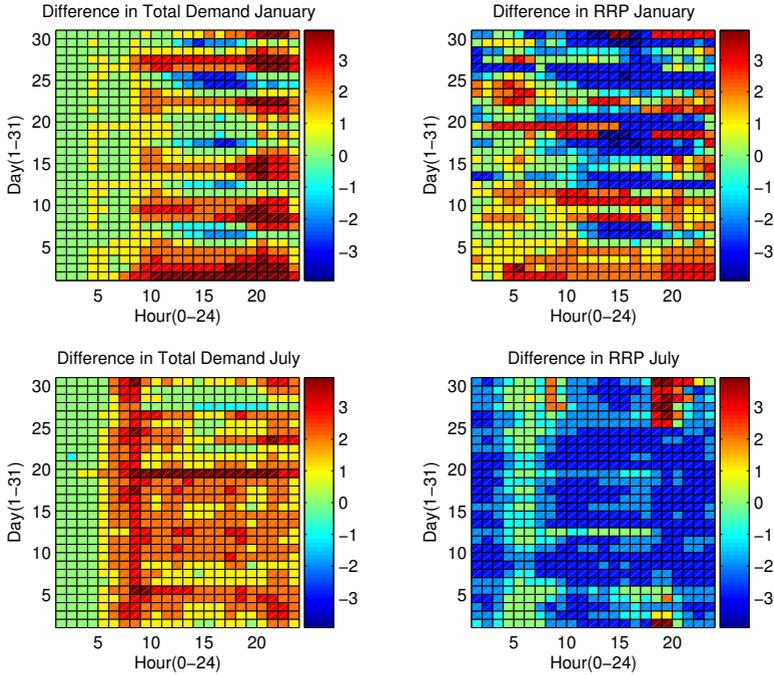


FIGURE 3: The Difference in Total Demand and RRP data for Queensland against the other 3 regions for January and July 2003.

formly close to that of the selected states. However, this second observation is not as pronounced as the first and needs further investigation. Horizontal rotation of the plots may improve the identification of these characteristics, but this is not possible in the current printed form. So, whereas this type of plot is visually interesting, much fine detail may be hidden. To investigate both of the characteristics, two further visualisation techniques were used.

The next technique applied to the data was the development of a discrete contour plot for each month. Then, for $i = 1, \dots, 12$ and $s_i \in \{28, 29, 30, 31\}$, the data in the $s_i \times 24$ array m_i was projected onto an $s_i \times 24$ contour plot.

Figure 3 shows this contour plot for the months of January and July 2003, for both TD and RRP data. The homogeneous nature of the first five hours of the TD data is clearly visible, as are the troughs, see also [5]. These plots all indicate that, from midnight to five a.m. each day, Queensland's TD was close to the average of the other selected states. However, the plots also indicate that for the summer months this characteristic stretches from midnight to eight or nine a.m., which may correlate with an increase in daylight hours in summer. But note that this characteristic is not exhibited in the evening. Interestingly, the RRP plot for January is unlike its TD counterpart. The TD and RRP plots for July show a little more similarity, but no obvious direct correlation between the two is apparent. As for characteristics across days of the week, while the contour plots show troughs they do not show regular characteristics corresponding to particular days of the week.

Another interesting fact arising from this visualisation, particularly for July, is that for most of the time Queensland's RRP is below the average of the three regions. As we get to the latter stages of 2003 this characteristic is more pronounced.

To investigate further the correlation between the TD and RRP for Queensland and the average TD and RRP for the other selected states we include a standard x, y scatter plot, see Figure 4. These graphs are based on the raw data, but with significant outliers being removed from the RRP data.

The correlation coefficient for the entire data sets was relatively high at $R^2 = 0.8979$ for the TD and $R^2 = 0.8760$ for the RRP. However, see that if we take a moving window of eight hours, starting at midnight and incrementing by one hour, the correlation coefficient decreases though the middle part of the day and into the evening. These values are displayed in Table 2 and indicate that there is a stronger correlation between Queensland and the average of the selected other states, for the early part of the day. This supports the conclusions made above.

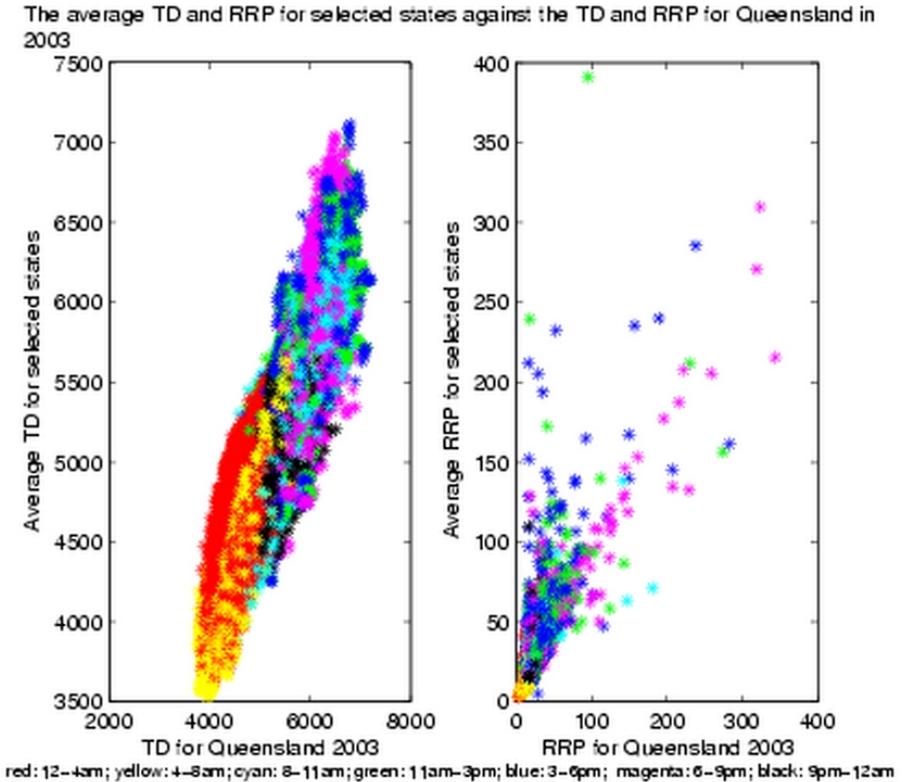


FIGURE 4: Data for selected states view against data for Queensland.

5 A final visualisation

The final visualisation technique employed here is a cumulative area function, which is plotted for the monthly matrices. Figure 5 displays this plot for TD and RRP data for the months of January and July 2003. Similar plots for each of the months may be viewed at [5].

In the cumulative area plot the x -axis represents hours and the value for

TABLE 2: The correlation coefficients for twenty four 8 hour windows.

From	12 am	1 am	2 am	3 am	4 am	5 am	6 am	7 am
R^2 for TD	0.85	0.90	0.93	0.95	0.95	0.93	0.89	0.84
R^2 for RRP	0.97	0.97	0.83	0.58	0.51	0.46	0.60	0.56
From	8 am	9 am	10 am	11 am	12 pm	1 pm	2 pm	3 pm
R^2 for TD	0.81	0.79	0.78	0.77	0.74	0.73	0.71	0.72
R^2 for RRP	0.57	0.48	0.53	0.89	0.89	0.90	0.90	0.90
From	4 pm	5 pm	6 pm	7 pm	8 pm	9 pm	10 pm	11 pm
R^2 for TD	0.73	0.75	0.78	0.82	0.85	0.85	0.83	0.82
R^2 for RRP	0.90	0.92	0.920	.91	0.96	0.97	0.97	0.98

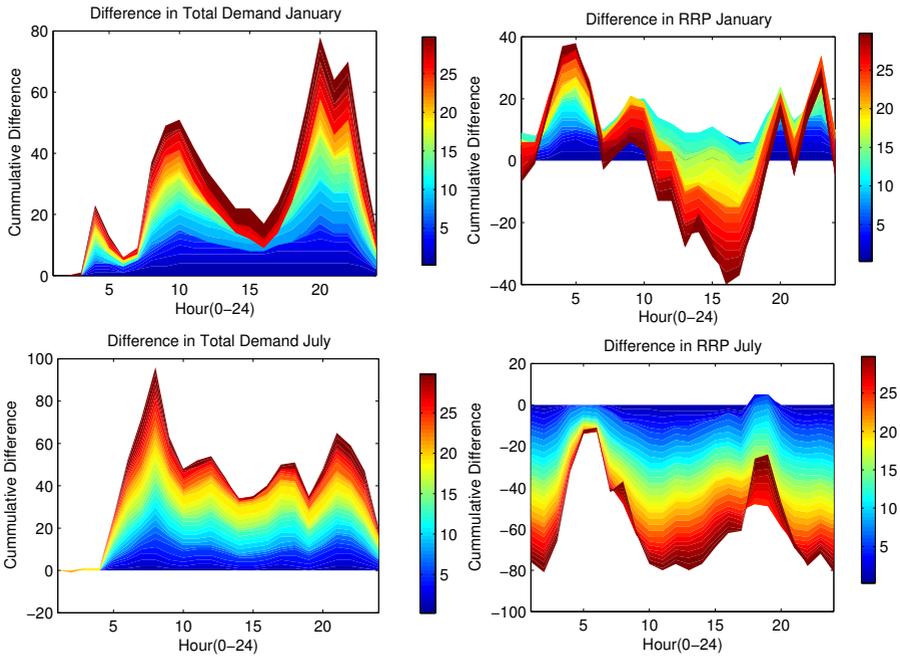


FIGURE 5: A cumulative area plot for Total Demand and RRP data for Queensland against the other 3 regions for January and July 2003.

any hour i is defined as

$$\sum_{\text{each day of the month}} \text{Difference in total demand (RRP) data for hour } i$$

So for a given time slot, hour i , the graph displays the progressive sum of the TD or RRP from the first day of the month to the last. The colour indicates the day of the month. If for a given time slot, the value of the difference in the TD, or RRP, changes sign from one day to the next, then the sum has a corresponding increase or decrease and so the graph laps back across itself. This visualisation technique again highlights the homogeneous nature of the TD data for the early part of each day. As these visualisations are viewed across the twelve months see that the RRP for Queensland does tend to be closer to the average RRP for the three states for hours 3 a.m. to 8 a.m. Such a characteristic is not as pronounced in the other two visualisations. Also the tendency of Queensland RRP data to fall below the average of the three regions is far more visible with this type of plot. Finally, the cumulative area function viewed across all months of the year suggests that there may be some correlation between Total Demand value at about 7:00 a.m. and 4:00 p.m., and the RRP value at these times. The plots for TD and RRP regularly show peaks and/or troughs around these times. For instance, July's Total Demand data has a spike at about 7:30 a.m. and the July's RRP data spikes just prior to this time.

6 Conclusion

Each of the above visualisation techniques revealed different information about characteristics in the data. The surface plots suggested a regularly occurring period in each day for which the TD and RRP data is close to the average of the three regions. This statement is supported by statistical analysis of the data. There is a hint of troughs occurring in the surface plot corresponding to complete days of little difference in the data. The contour

plots support this supposition, suggesting that this characteristic is affected by the month of the year and further suggests a slight correlation between TD data and RRP data. However, the contour plot does not provide evidence of a regular trend on a weekly basis. Finally, the cumulative area plots suggests a better correlation between these two types of data.

This research has suggests many new directions for future work on this problem, including a more rigorous statistical analysis of the data.

In conclusion, each visualisation technique has something interesting to add to the data analysis. Data plots of each type for all months of 2003 may be viewed at [5].

Acknowledgment: We acknowledge the assistance of Graeme Chandler and Pamela Burrage in regards to the MATLAB instructions.

References

- [1] Brown, R., and Pham, B., *Visualisation of Fuzzy Decision Support Information: A case Study* Fuzzy Systems, 2003. FUZZ '03. The 12th IEEE International Conference on Volume 1, pages 601–606, (2003). **C958**
- [2] Dwyer, T., and Eades, P. *Visualising a fund manager flow graph with columns and worms*. Proceedings of the Sixth International Conference on Information Visualisation (IV'02), (2002). **C960**
- [3] Keim, D., *Information Visualization and Visual Data Mining*, IEEE transactions on Visualization and Computer Graphics, Vol 7, No 1, pages 100–107, (2002). **C958**
- [4] Nemco Data [Online] <http://www.nemmco.com.au> (2004). **C959**

- [5] Visualising Energy Data [Online]
<http://www.maths.uq.edu.au/~dmd/energy/energydata.htm>
(2004). C963, C965, C966, C969
- [6] Unwin, A. If you can't see the pattern, is it there?. In D. J. Hand et al., editors *Pattern Detection and Discovery*, LNAI 2447, pages 63–76, Springer–Verlag (2002). C958